

MICROCOPY RESOLUTION TEST CHART NATIONAL BRIDGE OF COMPUTATION OF A

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT SUUNCLASSIFIED/UNLIMITED SAME AS RPT. DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
228 NAME OF RESPONSIBLE INDIVIDUAL R.W. PRESTRIDGE	22b. TELEPHONE (Include Area Code) 22c. OFFICE SYMBOL 512/482-5532

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

OFFICE OF NAVAL RESEARCH

Final Progress Report for the Period January 1, 1987 to December 31, 1987 Contract No. N00014-87-K-0180

POSITION FINDING AND TOLERANCES

D. B. Owen, Principal Investigator

Department of Statistical Science Southern Methodist University Dallas, Texas 75275-0332 Telephone No. (214) 692-2443

THE PROPERTY PROPERTY OF THE P

Reproduction in whole, or in part, is permitted for any purpose of the United States Government.

ACCOMPLISHMENTS

The results of the work on this contract are recorded in the attached paper which is scheduled for publication in January, 1988, in Communications In Statistics.

Accesion For	
NTIS CRASI VI	
DTIC TAB	
Upganounced [7]	
Jostifization	3
	COPY
The second secon	SPECTED
Ditt to 1	
Manually Comb	1
. ∮	···
1 m	;
	•
A_1	
Λ'	

ON THE DUALITY BETWEEN POINTS AND LINES

D.B. Owen

SEES PRODUCE COURSE COURSE STORES SYNONE ACCORDED

Jyh-Jen Horng Shiau

Department of Statistical Science Southern Methodist University Dallas, Texas 75275 Department of Statistics University of Missouri Columbia, Missouri 65211

Key Words and Phrases: Regression; directional data, normally distributed errors.

ABSTRACT

Duality between points and lines is examined, and the error distributions corresponding to various forms of duality are simulated. Several different least squares procedures for fitting of a line to several points are dualized to the fitting of a point to several lines and then dualized again to give the coordinate of a point of intersection of several lines.

1. INTRODUCTION

Duality appears in the statistical literature in several different forms, but none of these forms appear to be the same as here. For example, the term duality is used for the interchange of the role of X and Y in multiple correlation problems by Khatri (1964), who notes that Bartlett (1939) had first used this terminology. Duality is used extensively in linear programming. This type of duality has been exploited by Narula and Wellington (1977a and b) to obtain the minimum sum of weighted absolute errors in regression. It was also exploited by Pinski and Sposito (1976) and by Adriano (1977) to prove that normal equations are consistent. It is used by Book (1982) for least absolute deviations position finding.

Armstrong, Elam and Hultz (1977) use duality when dealing with a two-way classification model. Mardía (1972) has a book on

the statistics of directional data which deals with some of the topics considered in this paper, but in a different manner. The von Mises distribution, which is used for errors around a circle, is tabulated by Gumbel, Greenwood and Durand (1953). Still another form of duality exists in statistical designs and has an extensive literature. See, for example, Sinha and Dey (1983). Another form of duality exists in sample survey work. See Sirken and Casady (1982), or Lepkowski and Groves (1984), or Srivenkataramana (1980). Noorbaloochi and Meeden (1983) dualize unbiasedness and Bayes.

THE PROPERTY SHAPPED WINDSHIP BELLEVIES REPRESENTED

Duality of points and lines will be treated in this paper as problems in which lines are represented by coordinates and points are represented by equations. We will focus on problems of position finding, starting with those considered by Daniels (1951), Beale (1961), Rosenblatt (1978), and Hsu (1979). A recent paper by Souvaine and Steele (1987) exploits duality between points and lines and refers to some earlier work on the subject which was not known to the authors of this paper when this was written.

2. DUALITY BETWEEN LINES AND POINTS

Consider the equation of a straight line in the form y = mx + b, in the (X,Y) plane. We could just as well represent this line in a (U,V) plane by the coordinates (-m, b). Similarly, we could think of the point (x_1, y_1) in the (X,Y) plane and write an equation for it of the form $v = -x_1u + y_1$ in the (U,V) plane. Hence, there is a one-to-one correspondence between points and lines in the (X,Y) plane and lines and points in the dual (U,V) plane. Note that the common point of intersection of several lines $y = m_1x + b_1$ in the (X,Y) plane corresponds to the common line connecting the points $(-m_1,b_1)$ in the (U,V) plane.

Now consider the usual model in regression where $y = mx + b + \epsilon$ and ϵ is normally distributed with mean zero and variance \mathfrak{d}^2 . We have a sample (x_1,y_1) , (x_2,y_2) , ..., (x_n,y_n) and b and m are estimated by

$$\hat{m} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b = \overline{y} - m\overline{x}$$

CONTRACTOR DISCONDER PRODUCT (CONSIST PRODUCT CONTRACTOR DOCUMENTS)

The estimator of the regression of Y on X is $y = \hat{m}x + b$. Suppose now that we are interested in obtaining an estimator of the common point of intersection of several lines $y = m_i x + b_i$, $i=1, \ldots, n$. The equations of the lines are transformed into coordinates as outlined above, so that our data consist of (u_1, v_1) , (u_2, v_2) , ..., (u_n, v_n) where $u_i = -m_i$ and $v_i = b_i$.

We invoke the usual procedures to compute the regression in the (U,V) plane and obtain $v = \hat{m}u + b$, where \hat{m} and b are "the same as" \hat{m} and b above except that U's and V's are substituted for the X's and Y's. Then transforming this line back to the (X,Y) plane we have the coordinates of the estimated common point of intersection to be $(-\hat{m}, \hat{b})$. (Note the change in sign on the \hat{m} .)

The question arises about the error structure of this procedure. In the (X,Y) plane we assume the form $y_i = mx_i + b + \epsilon_i$ for $i = 1, 2, \ldots$ n where the ϵ_i s are independent. Hence, if we use the same formulas in the (U,V) plane we assume that $v_i = mu_i + b + \epsilon_i$ for $i = 1, 2, \ldots$, n where again the ϵ_i 's are independent realizations of ϵ which are normally distributed with mean 0 variance \mathfrak{o}^2 . An estimator of \mathfrak{o}^2 is

$$\int_{0}^{A} e^{2} = \frac{1}{n-2} \int_{i=1}^{n} (v_{i} - Mu_{i} - b)^{2}.$$

We know that (\hat{m}, \hat{b}) are jointly bivariate normal with means (m,b) and variance-covariance matrix

$$\begin{pmatrix}
\frac{\sigma^2}{\Sigma(u_i - \bar{u})^2} & \frac{-\sigma^2 \bar{u}}{\Sigma(u_i - \bar{u})^2} \\
\frac{-\sigma^2 \bar{u}}{\Sigma(u_i - \bar{u})^2} & \frac{\sigma^2 \Sigma u_i^2}{n\Sigma(u_i - \bar{u})^2}
\end{pmatrix}$$

Hence, we can construct confidence intervals for m and b (the coordinates of the point of intersection of n lines) as

where

COSCOSO CACALLA DESERVADO DE CONTRACTO DE

$$s = \int_{0}^{2} = \sqrt{\frac{1}{n-2}} \sum_{i=1}^{n} (v_i - mu_i - b)^2.$$

$$a_{00} = \frac{\sum u_i^2}{n\sum (u_i - \bar{u})^2}, \text{ and}$$

$$a_{11} = \frac{1}{\sum (u_i - \bar{u})^2},$$

and where $t_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the Student t-distribution with n-2 degrees of freedom.

These could then be used to give us a confidence region in the (X,Y) plane for the point of intersection we seek. This confidence region would be

$$\left(\frac{\mathbf{x} + \frac{\mathbf{A}}{\mathbf{m}}}{\sqrt{\mathbf{a}_{11}}}\right)^{2} - \left(2\frac{\mathbf{u} \cdot \mathbf{n}}{\sqrt{\mathbf{a}_{11}}}\right) \left(\frac{\mathbf{x} + \mathbf{m}}{\sqrt{\mathbf{a}_{11}}}\right) \left(\frac{\mathbf{y} - \mathbf{b}}{\sqrt{\mathbf{a}_{00}}}\right) + \left(\frac{\mathbf{y} - \mathbf{b}}{\sqrt{\mathbf{a}_{00}}}\right)^{2}$$

$$\leq 2\mathbf{s}^{2} \mathbf{F}_{2, \mathbf{n} - 2}^{(1 - \mathbf{\alpha})\sqrt{\mathbf{n}} \mathbf{a}_{00}}.$$

Hence, with probability $1-\alpha$ we can be sure that m and b lie in this region, where $F_{2,n-2}(1-\alpha)$ is the $(1-\alpha)$ th quantile of the F-distribution based on 2 degrees of freedom for the numerator and n-2 degrees of freedom for the denominator.

In the (U,V) plane we are, however, minimizing the sum of squares $\Sigma(v_i - \hat{n}u_i - \hat{b})^2$ with respect to m and b, i.e., we are assuming the u_i 's are obtained without error while the v_i 's contain all of the error. In the (X,Y) plane this means that the slope of each line is without error while the intercept, b, is associated with all of the error. That is, we are minimizing the sum of the squares of deviations measured parallel to the y-axis of each line from the point with coordinates (\hat{n},b) . Hence, the process we are using

minimizes $\sum_{i=1}^{n} (b-mu_i-v_i)^2$ with respect to b and m and gives

the solution (\hat{m}, b) . In the (X,Y) plane this corresponds to taking the vertical distance from the point to be found to each line and minimizing that sum of vertical distances squared.

For example, if we have six lines

Equation of line	Coordinates of line
y = 50.5	(0,50.5)
y = 46.8	(0,46.8)
y = 1.5x + 62.3	(-1.5, 62.3)
y = 1.5x + 67.7	(-1.5, 67.7)
y = 3x + 80.1	(-3 , 80.1)
y = 3x + 79.2	(-3 , 79.2)

Note that for this example the sample lines are parallel in pairs. Nevertheless, we want to estimate a common point of intersection.

Transforming to the (U,V) plane we get $\hat{m}=93/9=10.3333$ and b=48.9333, or in the (X,Y) plane the point of intersection is (-10.3333, 48.9333).

In the solution above, for a common intersection of n lines we are assuming that m is obtained without error for each of the lines, but that b (the intercept) is subject to error.

Therefore, in the (X,Y) plane we can be 95% sure that the point of intersection is inside the ellipse given by:

$$\left(\frac{x+10.33}{\frac{1}{3}}\right)^{2} - 2\left(\frac{1.5\sqrt{6}}{\sqrt{22.5}}\right)\left(\frac{x+10.33}{\frac{1}{3}}\right) \left(\frac{y-48.93}{\sqrt{22.5}}\right) + \left(\frac{y-48.93}{\sqrt{22.5}}\right)^{2}$$

$$\leq 2(6.9443)(5.6983)(0.4)$$

or

$$(x+10.33)^2 - 0.8(x+10.33)(y-48.93) + \frac{4}{15}(y-48.93)^2 \le 3.5174$$

or alternatively, we can have confidence intervals on m and b as follows. We can be 95% confident that m lies between

- 10.333
$$\stackrel{+}{=}$$
 2 776 (2.387) $\frac{1}{\sqrt{9}}$ or from -12.5421 to -8.1245

For b, we can be 95% confident that it lies between

48.933
$$\stackrel{+}{=}$$
 2.776 (2.387) $\frac{1}{\sqrt{9}} \frac{\sqrt{22.5}}{\sqrt{6}}$ or from 44.656 to 53.210.

3. SIMULATION RESULTS

To demonstrate the efficacy of the procedure in Section 2, we simulated the coverage of the above confidence intervals in two cases where the number of lines is (a) 30 and (b) 4. The

known point of intersection is (10,10) and $\emptyset = 1$. We computed 90% confidence intervals or regions. The tables give the percentage of actual coverage.

(a) Number of lines is 30:

# of Trials	d Seed	Slope(m)	<pre>Intercept(b)</pre>	<pre>Joint(m,b)</pre>
100	7373	93	95	88
100	35757	88	93	87
200	76373	93.5	88	89
200	87197	91.5	88.5	89.5

(b) Number of lines is 4:

# of Trials	<u>d Seed</u>	Slope(m)	<pre>Intercept(b)</pre>	Joint(m,h)
100	3517	87	89	88
100	3579	90	88	88
200	55373	93	88.5	91.5
20C	5579	90.5	89.5	88.5

These results agree with the 90% confidence target value. The d seed is the seed used in the IMSL routine GGNML which was used in this simulation to generate normally distributed errors $\{\epsilon_i\}$.

Figures 1 through 8 show a plot of the estimated points of intersection for the above data.

4. SECOND SOLUTION BASED ON DISTANCE FROM A POINT TO SEVERAL LINES

In Lindley (1947) the least squares fit for the equation y = mx + b based on the minimum sum of squares of distances from each point to the line is given by

 $\hat{m} = k - 1 / k^2 + 1$

FIGURE 1.

ESTIMATED INTERSECTION POINTS (x0, x0)=(10,10), SIGNA=1, PLINE=30 , IRIAL=100, DSELD=7373

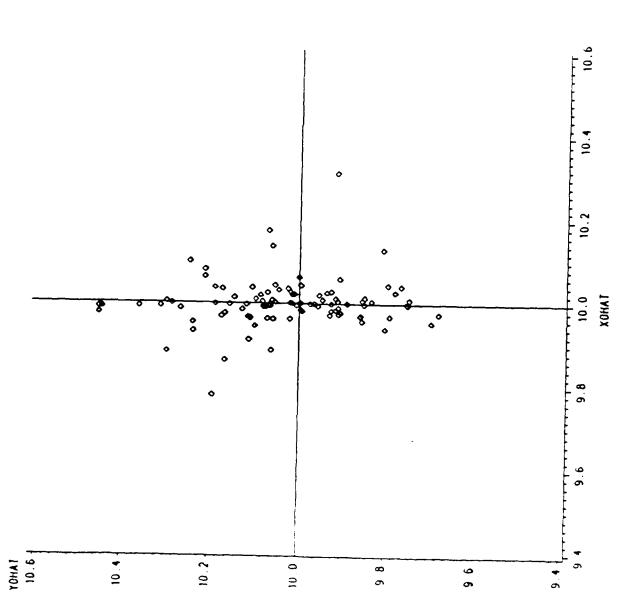
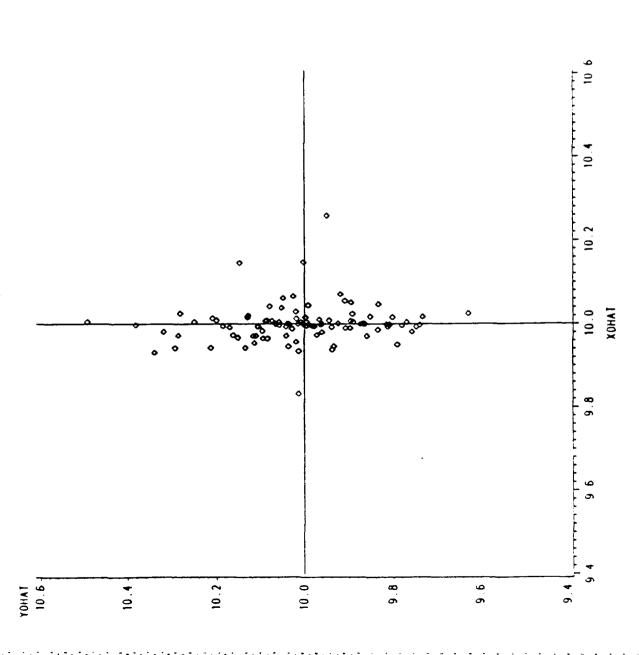


FIGURE 2.

F.S.TIMATED INTERSECTION POINTS (x0.Y0)=(10,10),SIGMA=1,ALINE=30 ,ITRIAL=100,DSEED=35757

KKK BIDIDA BIDISA RIZAKA BIDIDAK KKKKKK



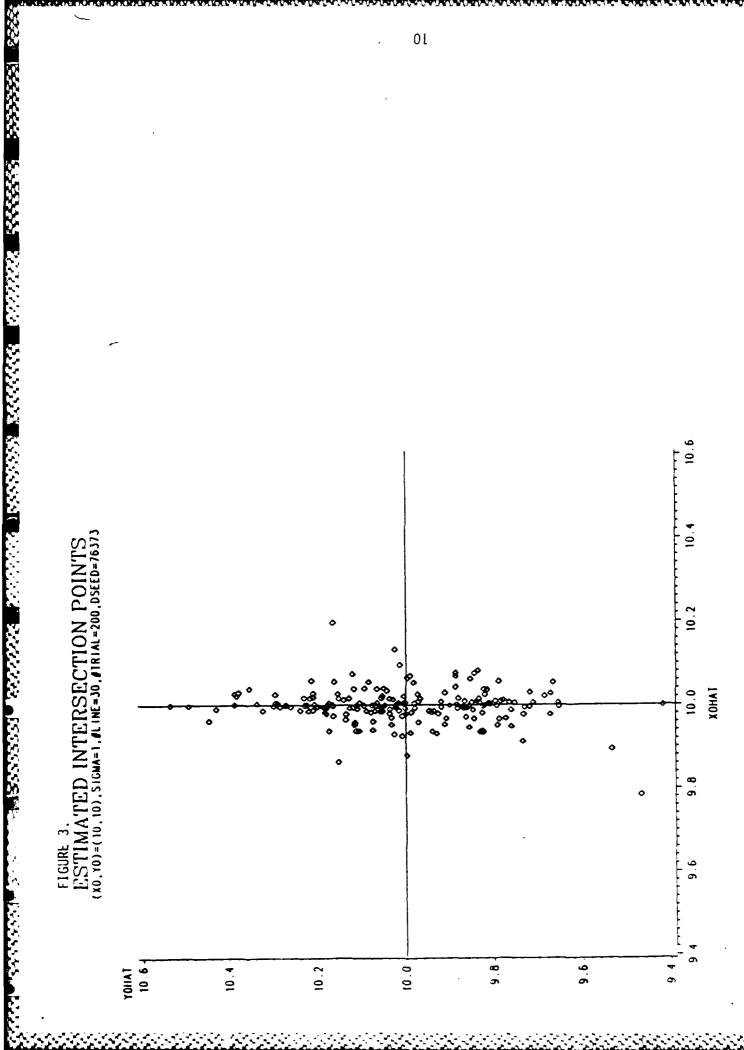
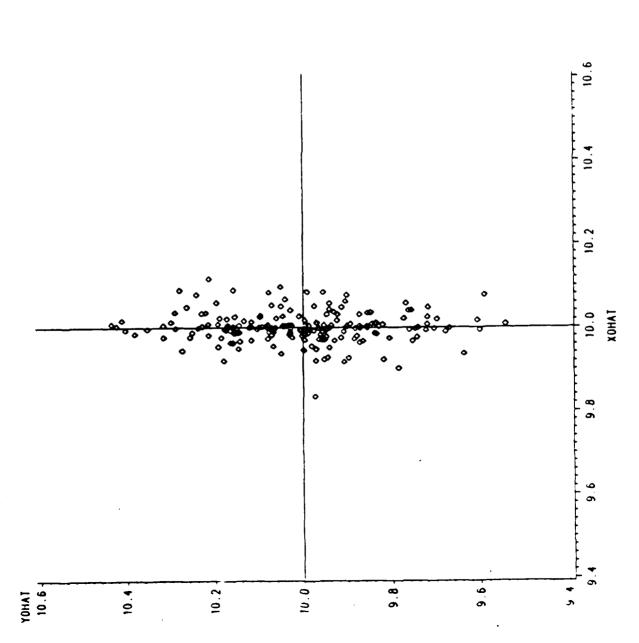


FIGURE 4.
ESTIMATED INTERSECTION POINTS (x0, Y0)=(10, 10), SIGMA=1. (LINE *30, 1) IRIAL = 200, DSEED= 87197

TARRA PROPERTY PROPERTY OF THE PROPERTY PROPERTY



ESTIMATED INTERSECTION POI

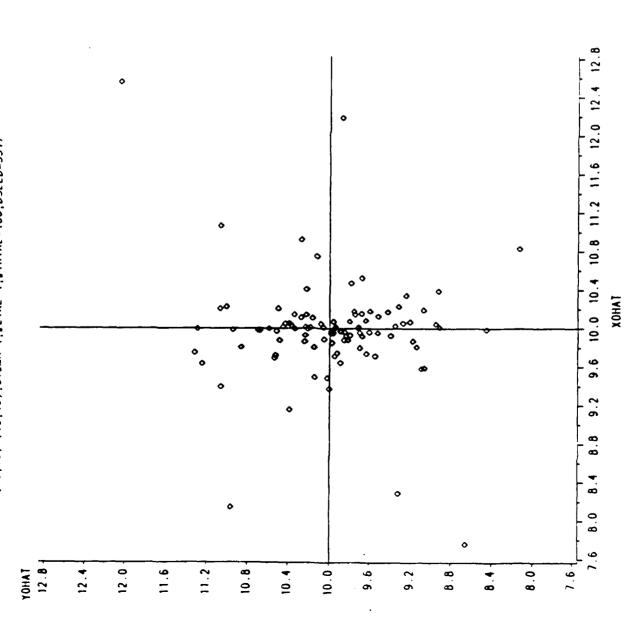
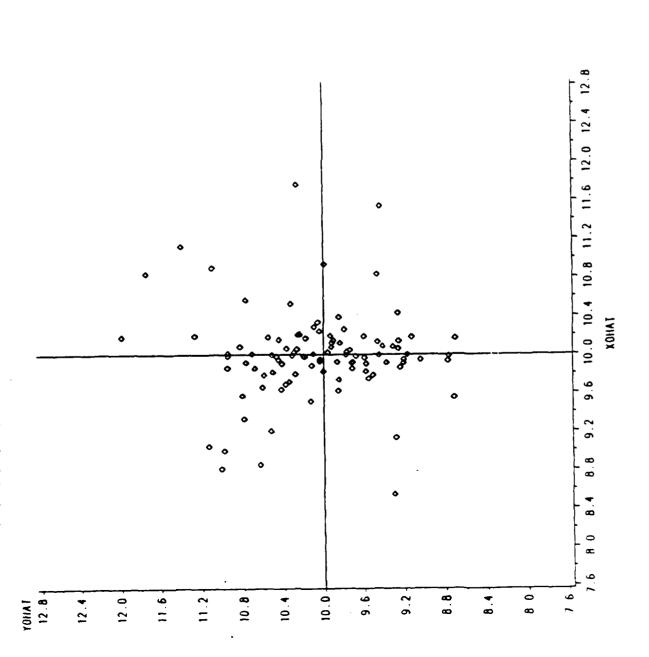


FIGURE 6.
F.S.T.IMATIED INTERSECTION POINTS (x0, x0)=(10,10), SIGMA=1, ILINE=4, ITRIAL=100, DSEED=3579

garral services escreve a receptor social social services. Receptor represent socialism receptor receptor



(3)

ESTIMATED INTERSECTION POINTS (x0, 70)=(10, 10), S1CM=), RINE *4, RINE *200, DSEED=55373

SEED BROKERS BROKERS SECTIONS WAYNESS PROPERTY OF THE PROPERTY INSTRUMENTS

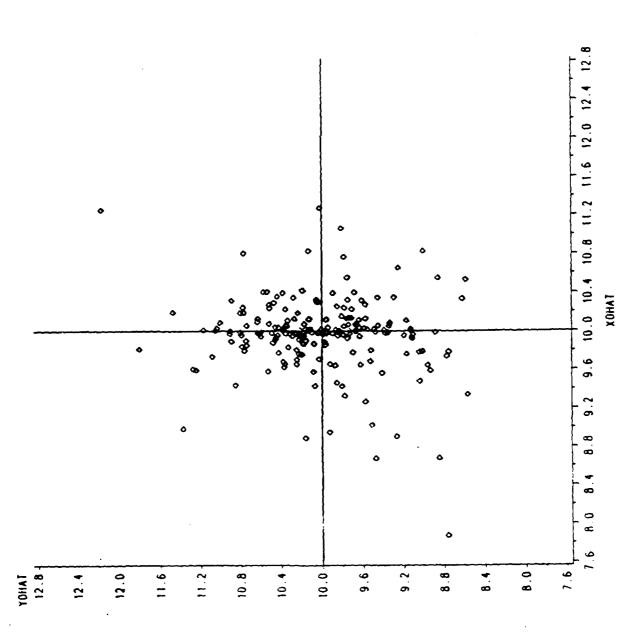
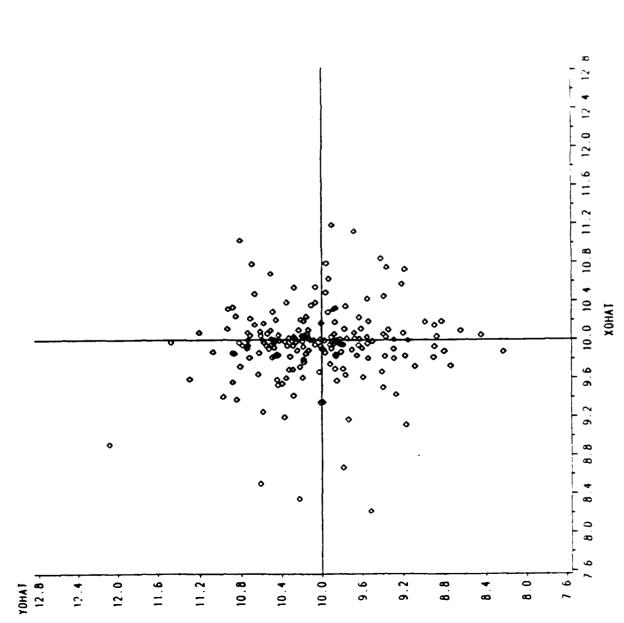


FIGURE 8.
ESTIMATED INTERSECTION POINTS (x0, x0)=(10, 10), SIGMA=1, #1 INE-4, #TRIAL=200, DSEED=5579



where $k = (SS_{yy} - SS_{xx})/2SS_{xy}$ and the sign of SS_{xy} is chosen for the $\frac{+}{}$ sign. This has been extended to the multivariate case by Fuller and Amemiya (1984). Here

$$SS_{yy} = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$$

$$SS_{xx} = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$$

$$SS_{xy} = \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}$$

Then b is found from $b = \bar{y} - \hat{m}\bar{x}$.

In the problem of Section 2

$$SS_{yy} = 983.79333$$

$$SS_{uu} = 9$$

$$SS_{uv} = -93$$
.

Hence,

$$k = -5.2408$$

or

$$\hat{m} = 10.5762$$

then

$$\hat{b} = \nabla - \hat{m} \hat{u}$$
.
= 64.4333 - 10.5762(1.5)
= 48.5690.

The coordinates of the point of intersection, assuming that sum of squares of distances from a point to the lines is minimized, become

$$(-10.5762, 48.5690)$$

in the (X,Y) plane, which should be compared with (-10.3333, 48.9333) which was obtained by ordinary least squares (OLS).

5. THIRD SOLUTION BASED ON PLUCKER LINE COORDINATES

Let us also consider the lines written in the form

$$\frac{mx}{\int_{m^2+1}^{2} - \frac{y}{\int_{m^2+1}^{2} + \frac{b}{\int_{m^2+1}^{2}}} = 0 .$$

It is in this form that the distance from a point to a line is calculated. Therefore, if the sum of squares of distances from a point to various lines is to be minimized we will assume the equation of a line is written in the form

$$ux + vy + 1 = 0$$

which is the form for the Plucker coordinates of projective geometry.

Then the Plucker coordinates for the lines in the example of Section 2 become

(0, -0.019802)

(0, -0.0213675)

(0.024077, -0.0160154)

(0.0221566, -0.014771)

(0.0374532, -0.0124844)

(0.0378788, -0.0126263).

The least squares line here is

$$v = 0.2129882u - 0.0204991$$

or

$$-10.390124u + 48.78262v + 1 = 0$$
.

Hence the intersection point is

(-10.3901, 48.7826).

6. INTERCHANGING ROLES OF X AND Y

For both the ordinary least squares solution and the Plucker line coordinate solution above, we can interchange the

role of X and Y and obtain two additional solutions. Note that, the second solution above is symetric in X and Y so the solution is the same.

In the case of Section 2, this amounts to the error being on the slope while the intercept is measured exactly. Since X and Y are merely interchanged, the roles of U and V are interchanged. More explicitly, the observed data consist of (u_1,v_1) , (u_2,v_2) , ..., (u_n,v_n) where u_i is the observed slope which is subject to error and v_i is the observed intercept which is now exact. Consider the model

$$u_i = c_1 v_i + c_0 + \epsilon_i$$
 $i = 1, 2, ..., n$

where the ε_i 's are independently and normally distributed with mean zero and variance \mathfrak{o}^2 . By interchanging the roles of X and Y, we simply mean that we regress $\{u_i\}$ on $\{v_i\}$ by the above model and obtain the regression line

$$u = c_1 v + c_0 ,$$

where \hat{c}_1 and \hat{c}_0 are OLS estimators of c_1 and c_0 , respectively. And then we can transform this regression line in the (U,V) plane back to the point in the (X,Y) plane, and obtain the point of intersection from

$$\left(\frac{1}{\hat{c}_1}, -\frac{\hat{c}_0}{\hat{c}_1}\right)$$
.

The points of intersection for the example of Section 2 are given in the following table for all the error structures.

Regression of Y on X		Regression of X on Y	
OLS:	(-10.3333, 48.9333)	(-10.5784, 48.5657)	
Fuller:	(-10.5762, 48.5690)	(-10.5762, 48.5690)	
Plucker:	(-10.3901, 48.7826)	(-10.7325, 48.3540)	

The error structure for the Plucker coordinates is difficult to characterize, but it is interesting to see the results in this example.

7. RELATIONSHIP TO OTHER APPROACHES

The problem considered here is directly related to work by Daniels (1951) and others and summerized in Mardia (1972). In this other approach position finding is characterized by several angular readings which follow a circular normal distribution. In Section 2 above, the error was on the intercept and in Section 6 (where X and Y were interchanged) the error was on the slope. We can think of the slope as $\tan\theta$ where θ is the angle a line makes with the horizontal axis. We might then ask what the distribution of θ looks like if $\tan\theta$ is normally distributed with mean zero and deviation one. This is a simple problem to solve but surprisingly we find that the distribution of θ for $-\pi/2 \le \theta \le \pi/2$ has two modes at $\frac{1}{\pi}\pi/4$.

In fact, the density looks almost uniform over the interval -1.25 $\leq \theta \leq$ 1.25. Data from this distribution would be hard to distinguish from data from a uniform distribution.

8. CONCLUSION

The ideas in this paper are easily extended to three and more dimensions. In fact, they may be even more practically useful in these higher dimensions where the alternative approach as given by Mardia (1972) and others has not been extended. If we think in terms of trying to locate objects in space or beneath the surface of the earth, (say, the focus of an earthquake), it is clear that the duality should be between planes intersecting the earth and points in space. This extension to higher dimensions will be considered in another paper.

9. BIBLIOGRAPHY

- Adriano, K. (1977), "Consistency of the normal equations via linear duality," <u>Communications in Statistics</u>, Vol. B6, pp. 97-99.
- Armstrong, R.D., Elam, J.J. and Hultz, J.W. (1977), "Obtaining least absolute value estimates for a two-way classification model," <u>Communications in Statistics</u>, Simulation and Computation, Vol. B6, 365-381.
- Bartlett, M.S. (1939), "A note on tests of significance in multivariate analysis," <u>Proceedings of the Cambridge</u> Philosophical Society, Volume 35, pp. 180-185.
- Beale, E.M.L. (1961), "Estimation of variances of position lines from fixes with unknown target position," <u>Journal of Research of the National Bureau of Standards</u>, Series D, Vol. 65, pp. 5-273.
- Book, S.A. (1982), "Least-absolute deviations position findings," <u>Naval Research Logistics Quarterly</u>, Volume 29, pp. 235-246.
- Daniels, H.E. (1951), "The theory of position findings,"

 Journal of the American Statistical Association, Vol. 48,
 pp. 131-152.
- Fuller, W.A. and Amemiya, Y. (1984), "Estimation for the multivariate errors-in-variables model with estimated covariance matrix," <u>Annals of Statistics</u>, Vol. 12, pp. 497-509.

- Gumbel, F.J., Greenwood J.A. and Durand, D. (1953), "The circular normal distribution: Theory and Tables," <u>Journal</u> of the American Statistical Association, Vol. 48, pp. 131-152.
- Hsu, D.A. (1979), "Long-tailed distributions for position errors in navigation," Applied Statistics, Vol. 28, pp. 62-72.
- Khatri, C.G. (1964), "Distribution of the 'generalized' multiple correlation matrix in the dual case," <u>Annals of Mathematical Statistics</u>, Vol. 35, pp. 1801-1806.
- Lepkowski, J.M. and Groves, R.M. (1984), "The impact of bias on dual frame survey design," American Statistical

 Association, Proceedings of Survey Research Methods, pp. 26⁵-270. D.V.

- Lindley, D.V. (1947), "Regression lines and the linear functional relationship," <u>Journal of the Royal Statistical Society</u>, Supplement 9, pp. 218-244.
- Mardia, K.V. (1972), <u>Statistics of Directional Data</u>, Academic Press, New York.
- Narula, S.C. and Wellington, J.F. (1977a), "An algorithm for the minimum sum of weighted absolute errors regression,"

 <u>Communications in Statistics</u>, Simulation and Computation, Vol. B6, pp. 341-352.
- Narula, S.C. and Wellington, J.F. (1977b), "Prediction, linear regression and the minimum sum of relative errors," Technometrics, Vol. 19, pp. 185-190.

- Noorbaloochi, S. and Meeden, G. (1983), "Unbiasedness as the dual Bayes," <u>Journal of the American Statistical</u>
 Association, Vol. 78, pp. 618-623.
- Pinski, K. and Sposito, V.A. (1976), "Consistency of the normal equations," <u>Communications in Statistics</u>, Vol. A5, pp. 1055-1057.
- Rosenblatt, J.R. (1978), "Statistical model for random errors of position location based on lines of position,"

 National Bureau of Standards, Internal Report #1457.
- Sinha, K. and Dey, A. (1983), "A series of truly self-dual PBIB designs," Mathematische Operationsforschung und Statistik, Vol. 14, pp. 53-54.
- Sirken, M.G. and Casady, R.J. (1982), "Nonresponse in dual frame surveys based on area/list and telephone frames,"

 American Statistical Association, Proceedings of Survey Research Methods, pp. 403-406.
- Souvaine, D.L. and Steele, J.M. (1987), "Time- and space efficient algorithms for least median of squares regression," <u>Journal of American Statistical</u> Association, Vol. 82, pp. 794-801.
- Srivenkataramana, T. (1980), "A dual to ratio estimation in sample surveys," Biometrika, Vol. 67, pp. 199-204.

END FILMED FEB. 1988 TIC